

---

# Report of the CSAC Differential Privacy Working Group- Task 4

Richelle Winkler (convenor) with  
Deborah Balk, Jay Breidt (ex officio), John Czajka,  
Barbara Entwisle, Kunal Talwar, Lance Waller

---

# Charter for CSAC Differential Privacy Working Group

Four tasks, with deliverables consisting of presentations at full CSAC meetings

- **Task 1.** Developing a summary of use cases.
- **Task 2.** Developing recommendations for prioritizing use cases for the administration of a "privacy-loss budget."
- **Task 3.** Developing metrics to assess the impact of differential privacy on the accuracy of decennial census data.
- **Task 4. Developing strategies for communicating the use of differential privacy for the 2020 Census data products.**

# Purpose

This presentation provides recommendations for messaging and communication strategies for informing data users, of all skill levels, about what differential privacy is, how the specific disclosure avoidance algorithms implemented in the 2020 DAS work, the data protection the 2020 DAS provides, the privacy/accuracy tradeoff, and the strengths and limitations of the 2020 Census data products.

# Trust & Transparency

*Public trust is at the core of disclosure avoidance strategies.*

Trust in privacy

Trust in the data

Transparency builds trust, but can also be a challenge

- Ignore noise when you don't know
- Initial demo data problems
- Public and legal discussion raising concerns about DP
- Users look at their own place and it seems “wrong” or “fake”
- Confusion on what data to expect and when
- Commercial developers advocating their data is better
- Census reputation could be at risk

# Draft Recommendations

- *Consult with stakeholders to identify common concerns about DP and develop concise and clear public responses.*
  - a. *Consider and respond to broader messaging that is currently “out there”*
- *Prepare and publish a detailed report on the public engagement and iterative process of developing the DAS system for Census 2020*
- *Provide a regularly updated roadmap and timeline for releases, noting what will be released when (including demonstration data)*
- *Create a forum (data user group) where users can ask/answer questions about differential privacy, similar to ACS Data Users Group*

# Communications Products

- Expert/academic engagement-- conferences, NASEM, demo data
- Public Engagement (still mostly expert target):
  - a. Videos- [Protecting Privacy with Math](#) (Sept 2019- 400K+ views);  
[Protecting Privacy in CB Statistics](#) (July 2021- 1K+ views)
  - b. Blog- [2020 Disclosure Avoidance System Updates](#) (March 2020+)
  - c. Webinar Series- [Disclosure Avoidance Webinar Series](#) (2021)
  - d. Handbook- Nov 2021**
    - i. [Disclosure Avoidance for the 2020 Census: An Introduction](#)- in partnership with the Population Reference Bureau.

# Specific Questions from CB- DAS Handbook

1. Is there any information missing from this handbook that would be useful?
2. Is any of the material too technical or confusing?
3. Should any of the content be reorganized or presented differently?
4. Were there any sections of the handbook that you found particularly useful? Any section that were less useful?

# General Comments

*WG commends CB and PRB for the work to produce this important handbook. It is a useful guide in explaining what DP is and how it works using clear and accessible language. It offers guidance and valuable practicalities for mid-skill users.*

- Handbook (and related communications) are core opportunities for building trust among data users
- Multiple audiences: expert, data users, general public. Handbook is trying to do it all.



# Q1- Is there any information missing from this handbook that would be useful?

- Clarify what information is at risk of disclosure
  - Some readers may interpret sections on the risk of data attacks to mean that highly sensitive information is at risk of disclosure. It is important to recognize what data Census 2020 collects.
- No census is perfect, there are always sources of error. Put differential privacy in perspective with other error sources.
- What are the limitations of DP?

# Q1- Is there any information missing from this handbook that would be useful?

Real world impact of DP on the data is critical. Sections addressing **fitness for use** are valuable and would benefit from further development & **more examples**, potentially tailored for different use cases.

- How do you assess fitness for custom geographies? More guidance is needed. This is a hard problem that users are going to need better guidance and tools.

# Q1- Is there any information missing from this handbook that would be useful?

Some places where more explanation is required-- specific editorial issues

- Tables in the appendix require text interpretation. What is the message readers should come away with?
- Discussion of generating negative population counts and adjusting to be non-negative-- how so? Are they made zero? Taken from elsewhere? "Issues are also resolved" (how?) + "...by....".
- P. 16 AIANNH example in Kansas not fully explained

# Draft Recommendations in Response to Q1

- *Handbook/future versions should recognize what information is collected in Census 2020 (e.g. age, race, sex, household relations), and so is at risk of disclosure.*
- *Limitations of DP and concise statement on the real world impact should be included.*
- *WG values discussions of fitness for use and related examples. These should be expanded to include discussion for how to assess fitness for use in custom geographies and to include more examples (based on use cases).*
  - a. *Develop a user tool based on a model (fed by 2010 demo data in comparison to 2010 SF1) that estimates uncertainty associated with any data point. Users can enter geography, variable, value and tool estimates level of uncertainty associated with DP.*

# Draft Recommendations in Response to Q1

- *Handbook and associated documents should be “living”, continuously updated on an accessible website. Updated based on feedback, new data releases (e.g. DHC), and use case examples.*
- *Commission a critical review to address audience usability from the perspective of diverse data users (skill levels) and offer feedback on detailed editing*
  - a. *Editorial review outside scope of CSAC committee -- some examples provided*
- *Handbook should review various sources of inaccuracy in 2020 and prior, to put differential privacy 2020 in perspective*

## Q2- Is any of the material too technical or confusing?

- Too technical for some in places, but other audiences need that.
  - Discussion of epsilon/rho allocation highly technical - point to a more detailed resource
- Confusing in places
  - Photo metaphor p. 7 (blurring applied incorrectly?)
  - Glossary
  - What does Epsilon 19.61 mean?
  - Be careful about statements that users might overgeneralize- pop thresholds for fitness
- Key questions & examples are addressed, but sometimes buried and these could be emphasized with call-out boxes, etc.
  - %s is a good example p. 5

# Draft Recommendations in Response to Q2

- *Think systematically about a variety of audiences with diverse skill/knowledge levels and how the Bureau can best engage with them in different outlets. This might include:*
  - *Consider the audience when presenting technical details, such material should be useful to readers how understand key concepts of DP but not thoroughly versed in its intricacies*
  - *Engaging with the press/responsive to journalist requests to reach a very general audience*
  - *Adding links to resources appropriate for different users/audiences*
  - *Creating different resources for different audiences*
  - *Creating one-page documents with specific guidelines for different data uses (common use cases)*
    - *Similar to p. 19 but for different kinds of users*
- *Include examples and key issues in call-out boxes/exhibits*

## Q3- Should any of the content be reorganized or presented differently?

- The document is too long for most users
  - DHC will be vastly more complex!
  - As written the technical appendix has small audience, limiting its usefulness. Text interpretation is needed.
- FAQs largely redundant with material presented earlier. Doesn't get to key concerns.



# Draft Recommendations in Response to Q3

- *Consider making a more general and concise version of the Handbook for a novice audience/data user*
- *FAQ section should be revised to avoid repetition and address expressed public concerns directly*

# Q4- Were there any sections of the handbook that you found particularly useful? Any sections that were less useful?

## Highly Useful

- Examples!
- Section 3 on Recommendations and Considerations where using data
- Box 1.1 is useful and good to come right at the beginning. May want to rename as “Executive Summary” or otherwise recognize it as such.
- Graphics in general, especially Box 1.1, Fig 1.3, Table 2.2, Fig 2.1

# Draft Recommendations in Response to Q4

*Continue to include several specific examples of use cases and the real world implications of DP as well as graphical representations throughout publications/communications.*

*The DHC will require many more (and different) examples.*

# Discussion